

Cooperativity in the annealing of DNA origamis.

Jean-Michel Arbona, Jean-Pierre Aimé and Juan Elezgaray.¹

¹*CBMN, UMR 5248, CNRS, 2 r. R. Escarpit, 33600 Pessac, France.*

Abstract

DNA based nanostructures built on a long single stranded DNA scaffold, known as DNA origamis, offer the possibility to organize various molecules at the nanometer scale in one pot experiments. The folding of the scaffold is guaranteed by the presence of short, single stranded DNA sequences (staples), that hold together separate regions of the scaffold. In this paper, we modelize the annealing-melting properties of these DNA constructions. The model captures important features such as the hysteresis between melting and annealing, as well as the dependence upon the topology of the scaffold. We show that cooperativity between staples is critical to quantitatively explain the folding process of DNA origamis.

PACS numbers: 87.15.Cc, 87.14.gk, 82.39.8j

I. INTRODUCTION

DNA origamis [1] are formed in an annealing process where a long single stranded DNA (ssDNA) scaffold extracted from a virus (M13mp18) is folded with the help of a set of ~ 200 short ssDNA strands, called staples. The resulting structure has a typical size of 100nm. Locally, an origami is made of double stranded DNA (dsDNA) helices, where one of the strands corresponds to the scaffold. Planar and 3D structures are obtained as the superposition of DNA helices, with basically only one constraint: the staples should run from one helix to the other in such a way that mechanical deformations are minimized. This seemingly simple process has opened the way to the fabrication of a variety of nanostructures in 2D or 3D, acting as biosensors [2] or nanorabots [3–5]. DNA origamis also provide with an easy tool to organize molecules such as proteins or carbon nanotubes [6] at the nanometer scale.

However, despite these innovative realisations, the folding process of DNA origamis remains poorly understood. Open questions include the exact role of the staple connectivity, the differences in yield for very similar origamis [1], or the factors that can influence the global stability of these nanostructures. From an experimental point of view, the process of formation of a DNA origami can be analysed by collecting the fluorescence intensity of a reporter dye [7] or by monitoring the variation of its UV absorption as a function of the temperature (melting or annealing curve for, respectively, increasing or decreasing temperatures) [8]. In this paper we study in depth the process of formation of DNA origamis through the analysis of melting and annealing curves obtained from UV absorption experiments.

The rest of the paper is organized as follows. Section 2 describes the formation of structures made of three ssDNA (called 'small origamis' in the following). Those can be considered as the building blocks of the usual DNA origamis. In this section, we also give a short description of the experimental techniques and the interpretation of UV absorption data. In sections 3 and 4, the model describing the energetic cost associated with the local connectivity of origamis, and the modelling of the annealing-melting processes are presented. Section 5 compares the melting curves of several DNA origamis with those predicted by the model. The final section presents some concluding remarks. A short report of this modelling [9] and a comparison between the model and experimental AFM data [10] have been published elsewhere.

II. THE BASIC FOLDING MECHANISM UNREVEALED BY SMALL ORIGAMIS.

The basic mechanism in the formation of a DNA origami is the hybridization of a staple to the scaffold. In this section we will consider very simple systems that help in the understanding of this basic process. In order to simplify the reasoning, let us first consider the situation where one staple (call it S_1) only binds to two separate parts of the scaffold (in Rothemund origamis, staples usually bind to three different parts of the scaffold). The binding to the scaffold necessarily proceeds in two steps. First, one of the two parts of the staple binds to the correct region of the scaffold: the zipping mechanism that ensures this binding is stopped by the existence of a noncomplementary region in the scaffold. At this stage, half of the staple is bound to the scaffold, forming a stiff double helix. The remaining unbound parts of the scaffold and the staple form flexible coiled coil structures. Fig. 1 illustrates the two possible situations that can be found. When the staple is in the 'outer position', the coiled coil structures are on the same side of the double helix. This is usually called a 'bulge' in the context of DNA thermodynamics [11]. On the other hand, when the staple is in the 'inner position', the coiled coil structures are on opposite sides of this double helix. Therefore, the probability that they merge is lower and the staple is expected to be much less stable in this position. In Rothemund's origamis, both situations (inner and outer) are found. Whereas the formation of a bulge (outer position) is well documented, it seems that the thermodynamics of the 'inner position' binding has not been considered previously. This is why we decided to study the small origami structures.

Another important aspect in the folding of DNA origamis is the interaction between staples. Following the simple reasoning of the previous paragraph, let us consider the influence of a second staple S_2 on the binding of S_1 . If the melting temperature of S_2 is higher than that of S_1 ($T_m(S_2) > T_m(S_1)$), by the time S_1 starts binding, with high probability both parts of S_2 are bound to the scaffold, reducing its effective elasticity and/or length. This could in turn facilitate the binding of the second part of S_1 , resulting in some kind of cooperative behaviour.

We designed a DNA construction (called small origami) made of two ssDNA 32b long (staples) and a 64b long ssDNA (scaffold). This structure is similar to DAO structures[12] and comparable in size and shape to JX and PX structures which have already been studied

experimentally [13], and theoretically [14], [15]. Three different sets of staples based on the same structure were chosen to quantitatively evidence cooperative effects during the binding of the staples. In the first two sets, the two staples (B1 and B2) have very different compositions: the sequence of B1 only contains A or T nucleotide whereas B2 only contains G or C nucleotide. Accordingly, their melting temperatures are far apart, respectively 57°C and 91°C . This allows to differentiate the two staples in the melting curve. The third set has two staples, B1m and B2m, designed with chemical sequences different enough to avoid mispairing with the 64b template B0. They have close melting temperatures (respectively 77°C and 80°C) as their AT/GC ratio are similar.

The topology of the binding is illustrated in (Fig. 1): each staple contains two contiguous parts, 32b long, that bind to the scaffold. In (Fig. 1a), B1 is in the 'outer' position, B2 is in the 'inner' position. It should be noted that besides the existence of an entropic hindrance, the inner position requires that double-helical domains stay in close contact, which could result in additional instability. Motivated by the previous considerations, we have investigated three different cases:

- B2 is located in the inner part, B1 in the outer part.
- B2 is located in the outer part, B1 in the inner part.
- B1m is located in the outer part, B2m in the inner part.

For each of these three cases, we have performed UV absorption measurements as a function of temperature. These measurements are based on the fact that hybridized bases absorb less than open bases: this is the so-called hyperchromic effect. The absorbance associated to any staple S_i is:

$$Abs_{S_i}(T) = (Abs_0(S_i) + Abs_0(\bar{S}_i)) (1 - h_{S_i} x_{S_i}(T)). \quad (1)$$

In this equation, $x_{S_i}(T)$ denotes the proportion of staples that are folded (at temperature T), $Abs_0(S_i)$ is the absorbance of S_i and $Abs_0(\bar{S}_i)$ the absorbance of its complementary staple, calculated according to [16]. We used the following relation for the hypochromicity h_{S_i} :

$$h_{S_i} = 0.287(1 - f_{gc}(S_i)) + 0.15f_{gc}(S_i) \quad (2)$$

where $f_{gc}(S_i)$ is the fraction of GC content in the staple S_i . This formula has been slightly modified from [16] to improve the fit against the experimental data for the small origamis. The degree of pairing $\theta(T)$ can be obtained from the raw absorbance measurements following the methods in [8].

In (Fig. 2a) we report the derivative of the melting curves $\theta(T)$ that show the behaviour of the B1 strand for different configurations. Fig. 2a.A corresponds to the melting curve of B1 with its complementary $\overline{B1}$: it shows a maximum peak at 57 °C and a half width of 4.5 K. We analyse first the case where B1 is outer. Without the staple B2 (Fig. 2a.B), the structure produces a loop or bulge, as described in [11] that introduces an entropic penalty and decreases the melting temperature to 48.5 °C and a half width of 6 K. Thus the folding is much less robust. When B1 and B2 are both present two events appear on the melting curve. The first one at 83 °C (not shown in Fig. 2a) corresponding to the folding of B2 in the inner position. Then Fig. 2a.C, B1 folds at a temperature higher than when it is alone, with a maximum peak at 51.5 °C but with a similar half width. Therefore, the inner staple B2 helps the pairing of the outer staple B1 by suppressing part of the entropic penalty related to the bulge formed by the scaffold. When B1 is in the inner part (Fig. 2a.D), its binding is significantly destabilized: the maximum peak is located at 41.5 °C, 15 K lower than the value of the dsDNA and with a half width as large as 12 K. This is three times larger than the one of $B1\overline{B1}$. Again, when the B2 staple is added (Fig. 2a.E), the pairing of B1 is stabilized with a maximum peak much higher located at 50 °C but still with a rather large half width of 8 K. These experimental results support the evidence of a strong correlation between the two strands: the presence of B2 helps the folding of B1 whatever its location. Moreover, these results also show that the location of the strand is of importance, the inner location being much less favourable.

Because B1 folds at a much lower temperature than B2, the binding of B2 is not influenced by the presence of B1. Fig. 2b illustrates the influence of the location (inner or outer) for B2. The double helix structure $B2\overline{B2}$ (Fig. 2b.A) has a maximum peak at 91°C with a half width of 3.5 K, while when B2 is in the inner position (Fig. 2b.B) the peak is located at 83 °C with a half width of 8 K. When B2 is in the outer position (Fig. 2b.C) the peak is located at 86.5°C and is slightly narrower (half width of 6.5 K). Therefore, the same differences between the inner and outer positions are observed whatever the chemical sequence involved.

The experimental results obtained with the B1m-B2m set of staples are shown in (Fig. 2c

and 2d). The same trends as in the B1-B2 case are observed but with less pronounced effects: the temperature shifts and the increase of the half width with respect to the melting curves of the double strands $B1m\overline{B1m}$ and $B2m\overline{B2m}$ are smaller. A correlation effect is also noticeable, and is now observed for both strands B1m and B2m. When the staple B1m is in the solution, the staple B2m, which folds at a higher temperature, shows a narrower peak at a maximum 1.5 °C higher than when it is alone. Therefore partial folding of the staple B1m helps the folding of the staple B2m. Similarly to what was observed in the previous case, the correlation effect is even more effective when we consider the influence of the B2m staple on the folding process of the B1m staple (Fig. 2c.D) with a shift of the melting temperature from 67 °C to 72 °C.

III. HYPOTHESIS FOR THE MODEL OF DNA ORIGAMI FOLDING.

The previous section considered the folding of simple DNA constructions that go beyond the simplest double stranded DNA structure. A visual inspection of Rothemund DNA origamis shows that small origamis can be considered as their building blocks. Therefore, we expect that the same physical mechanisms (cooperativity and topological effects) that explain the formation of small origamis will be present in the formation of larger origamis. However, we first need to make a few hypothesis in order to make the problem tractable. The basic difficulty is related to the huge number of possible configurations that need to be handled to compute average properties such as the number of open base pairs. Long linear structures of double stranded DNA can be computed rigorously because recurrence relations can be established in such cases [18],[17]. DNA origamis are highly connected structures bearing pseudoknots. This prevents the use of linear recurrences.

The DNA origami folding process can be generally described by a set of insertion reactions of the form

$$S_i + N(S_i) \rightleftharpoons N(S_i)S_i, \quad (3)$$

where S_i denotes the i -th staple, and $N(S_i)$ represents a particular set of neighbour staples (to be defined below) that influence the insertion of staple S_i . Eq. 3 will be made more precise in the following. To proceed further, we need to make further approximations.

Each staple S_i of length $|S_i|$ can be divided in parts that hybridize to non-contiguous regions of the scaffold. Let us note $S_i = S_{i,1} + S_{i,2} + \dots$ such a division of the strand sequence

(typically, each 32b staple is divided in three parts but other partitions are possible). Staples in the small origamis are divided in two parts: $S_i = S_{i,1} + S_{i,2}$.

Hypothesis 1: we will focus on configurations where each part $S_{i,k}$ is either completely hybridized to the scaffold or completely unbound. Moreover, we also disregard misfolded configurations, that is, staples that partially hybridize to the 'wrong' part of the scaffold. Notice that this assumption is plausible for the one layer origamis we consider here. For more complex, multi-layered structures, the staples are divided in smaller parts so that the probability to bind to the wrong part of the scaffold is considerably increased. We also disregard possible secondary structures of the scaffold.

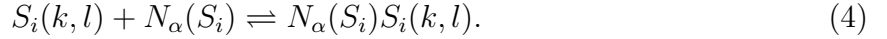
We will call *crossover* the connection between two contiguous parts of any staple. A crossover is not associated with a particular DNA base, it is only a convenient notation to describe the connectivity of the origami. In Rothmund's origamis [1], typical staples are 32 bases long and composed of three parts (8,16,8 bases long respectively) linked by two crossovers $cS_{i,1}$ and $cS_{i,2}$. On the scaffold side, a crossover is associated with a loop, a subset of the scaffold that is hybridized (or not) depending on the presence of other staples. In the previous example (Fig. 1a) B1 and B2 are composed of two parts connected by a crossover, and B0 plays the role of the scaffold.

Hypothesis 2: configurations with non contiguous hybridized parts are forbidden. This hypothesis is well verified when the central part of the staple is much longer than the other parts. In the following, we will note $S_i(k, l) (k \leq l)$ the configuration where $S_{i,k}, S_{i,k+1}, \dots, S_{i,l}$ are hybridized, the other parts being unpaired.

The model aims to compute the probability $p(S_i(k, l), T)$ of having a particular folded state of the staple S_i at temperature T . We will assume that at very high temperature $T = T_h$, all the staples are unfolded: $p(S_i(k, l), T_h) = 0$ (in practice, $T_h = 90^\circ C$). The model is *recursive*: $p(S_i(k, l), T + dT)$ is computed based on the knowledge of $p(S_i(k, l), T)$. The increment dT can be positive or negative: the algorithm starts from T_h , the temperature decreases down to a value T_l , then increases again. At any temperature T , the probability to observe a given configuration $S_i(k, l)$ will depend upon the presence (or not) of neighbour staples. Therefore, for each staple a set of possible neighbour staples $\{N_1(S_i), N_2(S_i), \dots\}$ needs to be defined. In the following, we will use the generic notation $N_\alpha(S_i)$ for any of these neighborhoods.

How many staples one has to consider in each of these sets is a parameter of the model.

In this paper, all the origamis are such that the crossovers can be aligned in rows. In the following, for each staple and crossover, the set of neighbour staples will be limited to those that have crossovers in the same row, and are separated by less than 75b. Let us note that all the results on this paper are robust against variations of this parameter. With these notations, the probability to observe the staple S_i in a given configuration $S_i(k, l)$ and for a given neighbourhood $N_j(S_i)$ is modeled by an equilibrium reaction:



This modelling therefore does not consider any kinetic effect. Notice however that the equilibrium hypothesis is a *local* one. Indeed, we expect that the model will reflect the hysteresis observed in the annealing-melting of DNA origamis.

Hypothesis 3: because the model only keeps track of the single probabilities $p(S_i(k, l), T)$ and not of the joint probabilities $p(S_1(k, l), S_2(k', l'), \dots, T)$, it is necessary to make an additional approximation to determine $p(N_\alpha(S_i), T)$. Again, based on the data from the small origamis, we assume that there is a strong correlation between the different staples. As the processes of annealing and melting are monotonous, for two staples S_{i_1} and S_{i_2} we make the hypothesis that if $p(S_{i_1}(k, l), T) < p(S_{i_2}(k', l'), T)$, then the S_{i_2} staple is present in the structure when S_{i_1} starts to fold. In order to compute $p(N_\alpha(S_i), T)$, let us generalize this idea and order the staples in $N_\alpha(S_i) = \{S_{i_1}, S_{i_2}, \dots\}$ in such a way that $p(S_{i_1}, T) \leq p(S_{i_2}, T) \leq \dots < p(S_c) = 1$, where S_c stands for the scaffold. According to the *high correlation hypothesis*, we approximate the joint probabilities in the following way:

$$\begin{aligned} p(S_{i_1}, T) &= p(S_{i_1}, S_{i_2}, \dots) \\ p(S_{i_2}, T) - p(S_{i_1}, T) &= p(S_{i_2}, S_{i_3}, \dots), \dots \end{aligned}$$

For instance, in the case where only two crossovers influence S_i , with probability $p(S_{i_1})$ both cS_{i_1} and cS_{i_2} are present, the probability of only having cS_{i_2} is $p(S_{i_2}) - p(S_{i_1})$ and the probability of only having the scaffold is $1 - p(S_{i_2})$.

The probability $p(S_i, T)$ for staple S_i to be hybridized at temperature T is the solution

of the coupled set of equations:

$$p(S_i, T) = \sum_{N_\alpha(S_i)} p(S_i, T | N_\alpha(S_i)) p(N_\alpha(S_i), T). \quad (5)$$

This set of equations can be further approximated by the iteration:

$$p(S_i, T) = \sum_{N_\alpha(S_i)} p(S_i, T | N_\alpha(S_i)) p(N_\alpha(S_i), T - dT). \quad (6)$$

Here, dT denotes the temperature step that determines the annealing-melting protocol and $p(S_i, T | N_\alpha(S_i))$ the conditional probability to hybridize staple S_i in the neighborhood $N_\alpha(S_i)$. In the annealing process ($dT < 0$), at high temperature T_∞ , $p(N_\alpha(S_i), T = T_\infty) = 0$ excepted for the empty neighborhood (the staple hybridizes to the scaffold in the absence of any other staple). For lower temperatures, $p(N_\alpha(S_i), T)$ can be computed from the knowledge of $p(S_i, T)$ and the correlations between staples (hypothesis 3). We show in appendix A that the set of coupled equations 5 determines $p(S_i(k, l), T)$ provided the equilibrium constants of the reactions 4 are known. This amounts to define an energy model which is detailed in the next section.

IV. MODELLING THE ENERGY OF DNA ORIGAMIS

In this section, we introduce an energy model for the hybridization of staples to the scaffold of a DNA origami. The gain in Gibbs free energy for hybridizing $S_i(k, l)$, $\Delta G(S_i(k, l), T) = \Delta H(S_i(k, l), T) - T\Delta S(S_i(k, l), T)$ contains two contributions:

$$\Delta G = \Delta G_{NN} + \Delta G_{top}. \quad (7)$$

The local contribution ΔG_{NN} only depends on the sequence of $S_i(k, l)$. It quantifies the gain in free energy associated with the local formation of double helices. We use the parameters of the nearest-neighbour model [11] with a temperature correction given in [19] (see appendix B and [19] for a detailed description).

ΔG_{top} gathers several contributions that depend on the connectivity of the origami (ΔG_{NN} depends mostly on the sequence of the scaffold and, to a less extent, on the density

of crossovers, but not on the connectivity). With any crossover, we associate an entropic penalty. This penalty reflects the difficulty for a staple to hybridize non-contiguous parts of the scaffold. In a first approximation, the longer the region of the scaffold that connects the two parts to be hybridized, the larger the penalty. Our previous results obtained with the small origami show that this needs to be refined. Based on these data, we consider three situations characterized by transient arrangements of staples that we call *local intermediate states* (LIS). In the first one (LIS1), the staple hybridizes to the scaffold, forming an internal asymmetric loop [11], Fig. 3a. The length of this loop corresponds to the number of unpaired bases of the scaffold linked by the crossover. This is a generalization of the 'outer' position found for the three strands origami. Before the crossover formation, when only part of the staple is folded, the scaffold and the non hybridized part of the staple are on the same side of the hybridized part of the staple (Fig. 1b). In this case, the staple is not involved in the path that connects the two extremities of the crossover. A particular case of LIS1, which we call LIS2, is the situation where the length of the loop is zero: the crossover forms locally a Holliday junction, the staple hybridizes in the close vicinity of an already hybridized staple (Fig. 3b).

The third LIS, LIS3, corresponds to the inner position in the small origami: the shortest path that connects the two ends of the crossover involves the staple itself (Fig. 3c). Because of this, before the crossover forms, the non hybridized parts of the strand and the scaffold are located on opposite sides of the hybridized parts (Fig. 1c). Therefore, LIS3 implies a larger penalty than LIS1 or LIS2. In the small origami, the shift in T_m was less than 10 °C for LIS1, between 5 °C and 7 °C for LIS2 and up to 15 °C for LIS3.

To each of these LIS is associated a different ΔG_{top} :

- $\Delta G_{top}(\text{LIS1}) = -T\Delta S_{bulge}(n_T - 0.8 \langle nb_{folded} \rangle)$. The function $\Delta S_{bulge}(n_T)$ is that of ref. [11]. n_T corresponds to the number of bases along the scaffold and $\langle nb_{folded} \rangle$ is the average value of bases folded along the scaffold (this average takes into account the probabilities of all the possible neighbouring configurations). The comparison between this model and the experimental results from the small origami structure is illustrated in Fig. 4.
- $\Delta G_{top}(\text{LIS2}) = \Delta H_{loop0} - T\Delta S_{loop0}$ with $\Delta H_{loop0} = 25.3 \text{ kcal/mol}$ and $\Delta S_{loop0} = 65.0 \text{ cal/mol/K}$. This constant contribution has been derived so as to fit as well as

possible the B1-B2 experimental data (left of Fig. 5) and then applied to the B1m-B2m data (right part of Fig. 5).

At the crossover, two bases that belong to S_i face the bases constituting the crossover made by the other strand. The initial enthalpic and entropic contribution of this pair of bases is subtracted from ΔG_{NN} as they are not nearest-neighbours anymore. One half of the contribution (nearest-neighbour model) of the two new pair of bases is added.

- $\Delta G_{top}(\text{LIS3}) = \Delta H_{unbind} - T\Delta S_{unbind} - T\Delta S_{bulge}(n_t - 0.8 < nb_{folded} >)$ (Fig. 3c): an additional penalty is added to the entropic penalty of the loop. This intends to reflect the stronger instability characteristic of this LIS. ΔH_{unbind} (resp ΔS_{unbind}) quantifies the loss of enthalpy (resp entropy) associated with the partial unfolding of the ends of an staple involved in this type of LIS. The number of bases that unfold is a parameter of the model. The data in Fig. 6 correspond to the unfolding of a total of 8 bases (two bases for each of the four extremities of the staple, see Fig. 3c).

Under some circumstances (Fig. 3d), two types of LIS can be attributed to a given crossover. In such cases, the LIS with the smaller ΔG_{top} is taken into account.

The modelling obtained with the contributions ΔG_{NN} and ΔG_{top} is quite satisfactory except for a constant negative shift ($\sim -4\text{K}$) of the melting temperatures. This shift indicates that another stabilizing mechanism that is not present in small constructions such as the small origamis has to be invoked. Indeed, in the folded structure of origamis, double-helix sections are separated by distances of the order of 1nm. It is then reasonable to think that mechanisms such as correlations between counter-ions and hydration forces come also into play, as is the case when DNA condensates [20]. This electrostatic stabilizing term is proportional to the number of neighbouring bases n_n that are close to the staple S_i , and to the length $|S_i(k, l)|$ of the partial configuration $S_i(k, l)$ considered.

$$\Delta H_n(S_i(k, l)) = -0.132n_n|S_i(k, l)|/|S_i|(kcal/mol) \quad (8)$$

The energy per base $-0.132kcal/mol/base$ is similar to the one needed for DNA condensation $10^{-1}k_B T/base$ to $10^{-2}k_B T/base$ [21]. An alternative explanation for the presence of this term would be the entropy reduction due to the confinement of the double helices in a 2D

structure.

The probability to fold any part $S_{i,j}$ of staple S_i is given by the sum of the probabilities of any configuration $S_i(k, l)$ that contains $S_{i,j}$:

$$p(S_{i,j}, T) = \sum_{l, p, l \leq j \leq p} p(S_i(l, p), T). \quad (9)$$

The fraction of folded bases for staple S_i

$$x_{S_i}(T) = \frac{1}{|S_i|} \sum_j p(S_{i,j}, T) |S_{i,j}| \quad (10)$$

can be then converted to a theoretical absorbance [16] as in the case of small origamis 1.

V. COMPARISON WITH THE ANNEALING-MELTING DATA OF DNA ORIGAMIS

We considered four different DNA origamis (Fig. 7) with the same scaffold (M13mp18 virus) and about 200 staples. O1 is the rectangle in the original Rothemund work [1], the staples are mostly 32b long, with folding part sequences divided in 8-16-8 patterns. O2 is another rectangular origami that includes a hole [23] and presents the same 8-16-8 pattern. O3 has the same connectivity pattern as O1, but some staples have been merged two by two in four areas (coloured in black in 7c), so that the typical staple pattern is 8-16-16-16-8. Finally, O4 is another rectangular origami where a 100b long subset of the scaffold goes from one side to the other of the rectangle, forming a ssDNA 'bridge'.

For each origami, a series of annealing-melting cycles was performed coupled to UV-absorption measurements. With the methods of [8], the degree of pairing $\theta(T)$ was extracted as a function of temperature. The temperature ramp ($0.4^\circ\text{C min}^{-1}$) is typical for this one-layer origamis. The annealing-melting process is not symmetrical, the hysteresis between the two phases of a cycle is such that the melting takes place at temperatures higher than the annealing.

There is an overall agreement (Fig. 7) between the melting-annealing curves observed experimentally and computed with the model. The model captures the hysteresis between the annealing and melting processes, as well as the relative strength of this hysteresis between

different origamis (O2 has only 4K shift between annealing-melting, whereas O4 has 10K shift). The maximum value of the derivative, which can be linked to the overall enthalpy of the transition in a two-state model, is also reproduced. This feature is robust against small variations of the parameters of the model.

A. Understanding DNA Origami design

In this section, we will rely on the model developed in the previous section to explore how the melting temperature of the rectangular origamis depends on the specific connectivity. This type of considerations could be relevant in applications where it is necessary to improve the stability of the template against temperature (DNA origamis as platforms controlling chemical reactions or other applications including grafting inorganic species).

1. *Circular permutation of the scaffold.*

The scaffold used for the design of rectangular origamis is a circular phage. Thus, it is possible to choose the beginning of the scaffold sequence anywhere so that 7248 different sets of staples (with same length and position, but with a different composition) are possible. We compared the melting curves given by those permutations on the O4 origami, permuting the sequence in steps of 16 bases as this shifts the middle of one staple to the middle of the nearby staple. The distribution of temperatures corresponding to the maximum of the derivative for the annealing and melting curves (Fig. 8a) shows an amplitude of variation of about 4°C . Depending on the permutation, the melting curves can be very different in shape (Fig. 8b). We conclude that, for a given scaffold, there is some sort of invariance regarding the choice of the origin.

2. *Decreasing the number of crossovers.*

In our model, a penalty is associated to each crossover. Reducing the number of crossovers should in principle increase the stability in the annealing process. We started from the O1 shape to reduce the number of crossovers, as its design is more regular. In the initial origami there is a length of 32 bases between two crossovers, which corresponds to three periods in

the double helix. Increasing the distance between crossovers leads to consider 54 bases (5 double-helix periods). We considered two possibilities, illustrated in Fig. 9: staples 27b long, split in 13-14 (O6 origami), and staples 54b long, split in 13-27-14 (O5 origami). Indeed, there is a trade-off between the gain in enthalpy when increasing the length of the staple, and the additional penalty of having two crossovers/ staple instead of only one. Our model shows that the net gain in stability, compared to the initial 8-16-8 staple strategy, is almost 20 °C (Fig. 9 O5). Again, the comparison with the experiments is excellent. Notice that decreasing the number of crossovers could have an impact on the flexibility of the origami.

VI. CONCLUSION

DNA origamis appear as a versatile tool to design various types of DNA based nanostructures. We have introduced a simple algorithm based on known thermodynamic properties of dsDNA and their parameterization with the NN-model [11]. The energy model is based on a preliminary study on small origamis that points to the importance of crossovers and their associated entropic penalty and correlation effects. This algorithm provides a reasonable account of the observed melting and annealing behaviour of DNA origamis. The model reproduces hysteresis and melting temperatures, as well as the width of the melting curve. The hypothesis of local equilibrium turns out to be compatible with the experimentally observed hysteresis. The model emphasizes the role of cooperativity in the folding process by introducing correlations between the probability of presence of neighbour staples. This shows that the folding of each staple is strongly dependent on the existence of a nearby cluster of folded staples. An intuitive picture can be derived from this modelling. In the annealing process, only the staples with a high GC content are able to fold at high temperature. These staples provide with initial nuclei that facilitate the folding of the neighbour staples. Therefore, the folding of staples with high AT content will take place at a temperature which is dependent on the distance to the set of these initial nuclei. During the melting process, the unfolding of any staple is mostly independent on its neighborhood and it is well accounted by an average penalty in free energy. The histeresis is then explained by the fact that the neighborhood of each staple is different in the annealing or melting.

As a possible application, we have shown that the model allows to improve the thermal stability by quantifying the effect of different construction factors such as staple length and

density of crossovers. Extensions to 3D [24] [25] and structures other than origami [26] are envisioned, as well as tests at the single molecule level (FRET). As shown in [10], the interaction with the depositing surface will be crucial to interpret future AFM measurements on the folding of DNA origamis.

Acknowledgments

This work was partially supported by CNRS (PIR funding). Thanks to Carmelo di Primo for letting us use its equipment. Thanks to Jean-Louis Mergny for letting us use its equipment and to Thao Tran for the Origami O2.

Appendix A: Computing the probabilities from the law of mass action.

In this section, we explain how to compute the probability $p(S_i(k, l), T)$ by solving a set of coupled equations which reflect the assumption that the reaction

$$S_i(k, l) + N_\alpha(S_i) \rightleftharpoons N_\alpha(S_i)S_i(k, l)$$

is governed by the law of mass action. For the sake of clarity, we consider first the particular case illustrated in Fig. 10: a staple divided in two parts of equal length is inserted in the vicinity of another staple that holds together a portion of the scaffold. For this particular example, we need to consider three equilibriums: the simultaneous binding of the two parts of S_i , with equilibrium constant:

$$K_{S_i(1,2)}^{N_\alpha(S_i)} = \exp \left(-\frac{\Delta G_{NN}(S_i(1, 2), T)}{kT} \exp(-\Delta S_{bulge}(2L))/k \right) \quad (\text{A1})$$

The binding of only one half $S_i(1, 1)$ of the staple, with equilibrium constant:

$$K_{S_i(1,1)}^{N_\alpha(S_i)} = \exp \left(-\frac{\Delta G_{NN}(S_i(1, 1), T)}{kT} \right) \quad (\text{A2})$$

The binding of the other half $S_i(2, 2)$ of the staple:

$$K_{S_i(2,2)}^{N_\alpha(S_i)} = \exp \left(-\frac{\Delta G_{NN}(S_i(2, 2), T)}{kT} \right) \quad (\text{A3})$$

More generally, for any set of equilibriums between S_i and its neighborhood $N_\alpha(S_i)$, the law of mass action reads:

$$\frac{[S_i(l, p)N_\alpha(S_i)]}{[N_\alpha(S_i)][S_i]} = K_{S_i(l, p)}^{N_\alpha(S_i)} = \exp\left(-\frac{\Delta G(S_i(l, p), T)}{kT}\right) \quad (\text{A4})$$

The concentration of free staples in solution is given by

$$C_0 \left(Exs - \sum_{\beta} \sum_{u, v, u \leq v} p(S_i(u, v) | N_\beta(S_i)) p(N_\beta(S_i)) \right)$$

Exs is the excess of staple concentration, compared to the initial concentration of scaffold C_0 , $p(S_i(u, v) | N_\beta(S_i))$ the conditional probability to observe $S_i(u, v)$ given the neighbour staples $N_\beta(S_i)$ and $p(N_\beta(S_i))$ the probability of the neighborhood N_β . In order to get a closed set of equations, we approximate $p(N_\beta(S_i), T) \sim p(N_\beta(S_i), T - dT)$ where dT is the temperature step in the melting-annealing process.

As the excess in DNA origami annealing experiments is important, the concentration of free staples will vary weakly in the process of formation. It is then approximated by $C_0 Exs$. The concentration of free N_α (S_i not bounded) is given by:

$$C_0 \left(p(N_\alpha(S_i)) - \sum_{u, v, u \leq v} p(S_i(u, v) | N_\alpha(S_i)) p(N_\alpha(S_i)) \right).$$

The concentration of the configuration $S_i(l, p)$ hybridized in the neighborhood $N_\alpha(S_i)$ is

$$C_0 p(S_i(l, p) | N_\alpha(S_i)) p(N_\alpha(S_i)).$$

The approximated expression of the equilibrium constant is therefore:

$$\begin{aligned} & K_{S_i(l, p)}^{N_\alpha(S_i)} \\ & \sim \frac{p(S_i(l, p) | N_\alpha(S_i)) p(N_\alpha(S_i))}{Exs C_0 \left(p(N_\alpha(S_i)) - \sum_{u, v, u \leq v} p(S_i(u, v) | N_\alpha(S_i)) p(N_\alpha(S_i)) \right)} \end{aligned} \quad (\text{A5})$$

This set of equations can be solved by noticing that the denominator in A5 is independent

of $S_i(l, p)$. Thus, for any couple of configurations $(S_i(l, p), S_i(u, v))$,

$$\frac{p(S_i(l, p)|N_\alpha(S_i))}{p(S_i(u, v)|N_\alpha(S_i))} = \frac{K_{S_i(l, p)}^{N_\alpha(S_i)}}{K_{S_i(u, v)}^{N_\alpha(S_i)}} \quad (\text{A6})$$

Let $y_\alpha = \sum_{l, p, l \leq p} p(S_i(l, p)|N_\alpha(S_i))$ and $S_\alpha = \sum_{l, p, l \leq p} K_{S_i(l, p)}^{N_\alpha(S_i)}$. Then, by summing over l and p eq. A5 reduces to:

$$\frac{1}{C_0 Exs} \frac{y_\alpha}{1 - y_\alpha} = S_\alpha \quad (\text{A7})$$

or:

$$y_\alpha = \frac{C_0 Exs S_\alpha}{1 + C_0 Exs S_\alpha} \quad (\text{A8})$$

and finally

$$p(S_i(l, p), T) = \sum_{\alpha} p(S_i(l, p)|N_\alpha) p(N_\alpha(S_i), T) \sim \sum_{\alpha} p(N_\alpha(S_i), T - dT) y_\alpha \frac{K_{S_i(l, p)}^{N_\alpha(S_i)}}{S_\alpha} \quad (\text{A9})$$

Appendix B: Temperature and salt corrections

The computation of the nearest-neighbour contribution $\Delta G_{NN} = \Delta H_{NN} - T \Delta S_{NN}$ includes corrections to take into account temperature and salt variations.

1. Temperature corrections

Besides the nearest-neighbour contributions $(\Delta H_N^0, \Delta S_N^0)$ [11], we have also included temperature dependent corrections:

$$\begin{aligned} \Delta H_{NN}(S_i(l, p)) &= \Delta H_N^0(S_i(l, p)) + C_p |l - p + 1| (T - T_{ref}) \\ \Delta S_{NN}(S_i(l, p)) &= \Delta S_N^0(S_i(l, p)) + C_p |l - p + 1| \ln\left(\frac{T}{T_{ref}}\right) \end{aligned}$$

where $C_p = -42 \text{ cal/mol/K/bases}$ and $T_{ref} = 53^\circ \text{C}$ according to [19].

2. Salt corrections

The parameters of the NN model [11] are given for standard salt concentrations ($[\text{Na}]=1\text{M}$, $[\text{Mg}]=0$). Different salt conditions can be taken into account using the correcting terms in [22]. These corrections apply when Mg is dominant. We assume that for each attachment of the staple S_i on the configuration $S_i(l, p)$ (of length $|S_i(l, p)|$), the relation

$$\frac{1}{T_m(Mg, Na, C, S_i(l, p))} = \frac{1}{T_m(0, Na, C, S_i(l, p))} + f(Mg, Na, f_{GC}, |S_i(l, p)|)$$

between the melting temperatures T_m at different concentrations of Mg and Na holds (f_{GC} is the fraction of GC content of $S_i(k, l)$).

The melting temperature corresponding to the point where $\Delta G(T_m) = 0$:

$$T_m(0, Na, C, S_i(l, p)) = \frac{\Delta H_{NN}(S_i(l, p))}{\Delta S_{NN}(S_i(l, p)) + R \ln(C/4)} \quad (\text{B1})$$

it can be deduced that the salt corrections are taken into account by an entropic correction given by:

$$SC(Mg, Na) = f(Mg, Na, f_{GC}, |S_i(l, p)|) \Delta H \quad (\text{B2})$$

where ΔH is the sum of all the contributions previously cited (nearest-neighbour, temperature corrections, topological contributions). In this paper, we modified the function $f(Mg, Na, f_{GC}, |S_i(l, p)|)$ as given in [22] by a small additive term to take into account low AT content strands. In order to do this, we compared the predictions of [22] with the experimental results obtained with the small origamis. In [22], the authors calibrated their model against 17 different dsDNA, involving a wide range of salt concentrations, and obtained a mean deviation -1.7 ± 0.7 °C . We deal here with a more restricted range of salt concentrations, it is therefore expected that the model [22] can be improved. We used a correction to the f function of [22] for the low fractions f_{GC} of GC content, to obtain a mean deviation of -0.4 ± 0.3 °C . The corrected f function reads:

$$f_{new} = f + (f_{gc} - 0.5)0.00008 \quad \text{if } f_{gc} < 0.5.$$

The results obtained with this new expression for the strands involved in the differents

small origamis are illustrated in Fig. 11.

-
- [1] P. Rothmund, *Nature* **440**, 297–302 (2006).
 - [2] Y. Ke, S. Lindsay, Y. Chang, Y. Liu, and H. Yan, *Science* **319**, 180 (2008) .
 - [3] K. Lund, A. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M. Stojanovic, N. Walter, et al. *Nature* **465**, 206–210 (2010).
 - [4] H. Gu, J. Chao, S. Xiao, and N. Seeman, *Nature Nanotech.* **4**, 245 (2009) .
 - [5] S. Wickham, M. Endo, Y. Katsuda, K. Hidaka, J. Bath, H. Sugiyama, and A. Turberfield *Nature Nanotech.* **6**, 166–169 (2011).
 - [6] H.T. Maune et al., *Nature Nanotech.* **5**, 61 (2010).
 - [7] C. Castro, F. Kilchherr, D. Kim, E. Shiao, T. Wauer, P. Wortmann, M. Bathe, and H. Dietz, *Nature Methods* **8**, 221–229 (2011).
 - [8] J.L. Mergny and L. Lacroix, *Oligonucleotides* **13**, 515–537 (2003).
 - [9] J.M. Arbona et al., *Europh. Lett.* (2012) submitted.
 - [10] J. Song, J.M. Arbona, Z. Zhang et al., *JACS* **134** 9844–9847 (2012).
 - [11] J. SantaLucia Jr, and D. Hicks, *Annual review of biophysics and biomolecular structure* **33**, 415 (2004).
 - [12] T. Fu and N. Seeman, *Biochemistry* **32**, 3211–3220 (1993).
 - [13] C. Spink, L. Ding, Q. Yang, R. Sheardy, and N. Seeman, *Biophys. J.* **97**, 528–538 (2009).
 - [14] P. Maiti, T. Pascal, N. Vaidehi, and W. Goddard III, *Nucleic acids research* **32**, 6047 (2004).
 - [15] P. Maiti, T. Pascal, N. Vaidehi, J. Heo, and W. Goddard, *Biophys. J.* **90**, 1463–1479 (2006).
 - [16] A. Tataurov, Y. You, and R. Owczarzy, *Biophys. Chem.* **133**, 66 (2008).
 - [17] D. Jost and R. Everaers, *Biophys. J.* **96**, 1056–1067 (2009).
 - [18] D. Poland and H. Scheraga, *Theory of helix-coil transitions in biopolymers.* (Acad. Press, 1970).
 - [19] C. Hughesman, R. Turner, and C. Haynes, *Biochemistry* **50** 2642 (2011).
 - [20] H. Strey, R. Podgornik, D. Rau, and V. Parsegian, *Curr. Opin. Struct. Biol.* **8**, 309–313 (1998).
 - [21] V. Bloomfield, *Biopolymers* **44**, 269–282 (1997).
 - [22] R. Owczarzy, B. Moreira, Y. You, M. Behlke, and J. Walder, *Biochemistry* **47**, 5336–5353 (2008).

- [23] M. Endo, Y. Katsuda, K. Hidaka, and H. Sugiyama, *J. Amer. Chem. Soc.* **132**, 1592–1597 (2010).
- [24] S. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf, and W. Shih, *Nature* **459**, 414 (2010).
- [25] H. Dietz, S. Douglas, and W. Shih, *Science* **325**, 725 (2009).
- [26] M. Hansen, A. Zhang, A. Rangnekar, K. Bompiani, J. Carter, K. Gothelf, and T. LaBean, *JACS* **132** 14481 (2010).
- [27] A. Rajendran, M. Endo, Y. Katsuda, K. Hidaka, and H. Sugiyama, *JACS* **133** 14488 (2011).

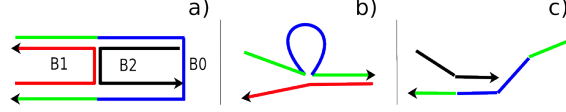


Figure 1: (a) Schematic representation of the connectivity of the small origami. (b) and (c) Transitory situation corresponding to the binding of half of the staple to the scaffold. (b) B1 staple is in the 'outer' position, (c) B2 staple in the 'inner' position. (b) and (c) show that the partial binding of staples in the outer (b) or inner (c) positions are very different.

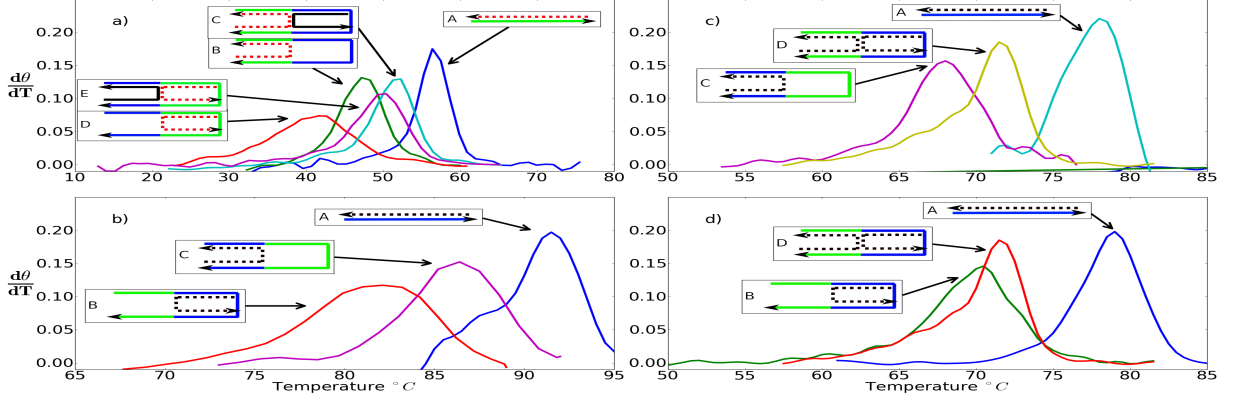


Figure 2: The derivative $d\theta/dT$ reported in the four figures corresponds to the folding of the dotted staple. a) experimental data on the folding of B1(AT) cases (A,B,D) in the absence of B2(GC), cases (C,E) with B2 already folded; b) experimental data on B2 without B1; c) experimental data on B1m; d) experimental data on B2m

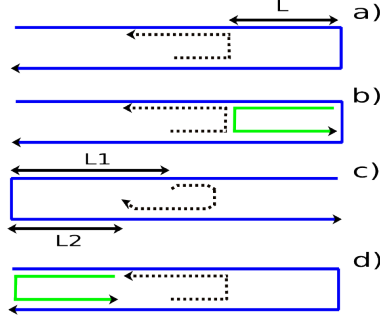


Figure 3: representation of the three different local intermediate states (LIS). The staple to be inserted is represented by the dotted line, the scaffold by the continuous line. (a) LIS1 corresponds to the formation of a bulge. (b) LIS2 corresponds to the case where, due to the binding of other staples, the effective length of the bulge is zero. (c) LIS3 generalizes the 'inner position' of the small origamis. We assume that, because of the curvature constraints imposed by this configuration, the staple is not able to fold completely. (d) A typical situation where two types of LIS (LIS1 at the right side of the staple, LIS3 at the left side) can be attributed to a given crossover.

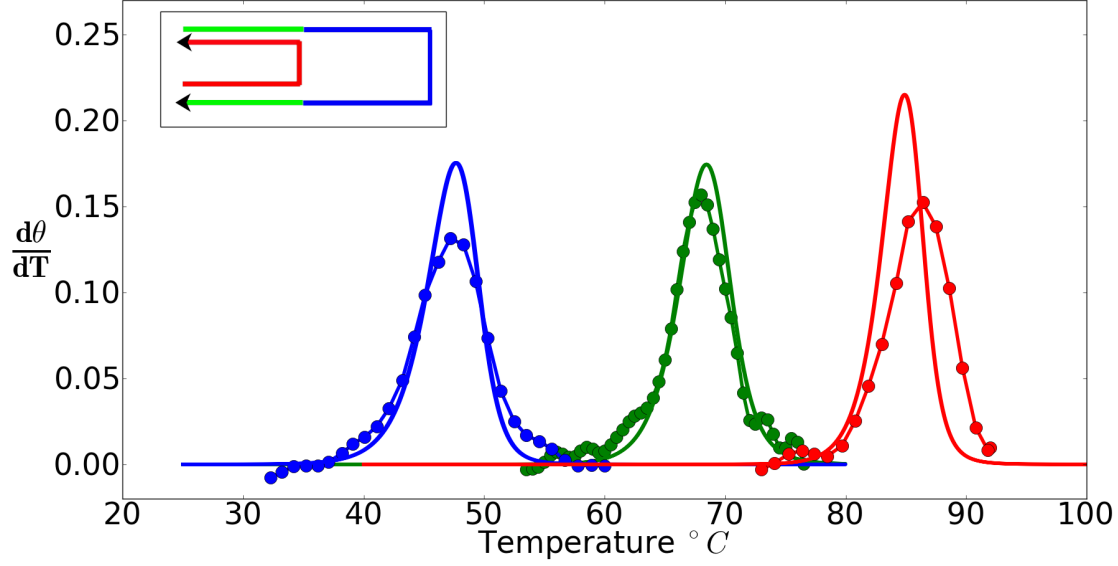


Figure 4: Melting curves obtained for a small origami containing only one staple in the outer position (inset). Comparison between the experimental (dotted line) derivative $d\theta/dT$ and the modeling (continuous line) as provided by the contribution $\Delta G_{top}(\text{LIS1})$. In order of increasing melting temperature: B1 staple (blue), B1m staple (green) and B2 staple (red).

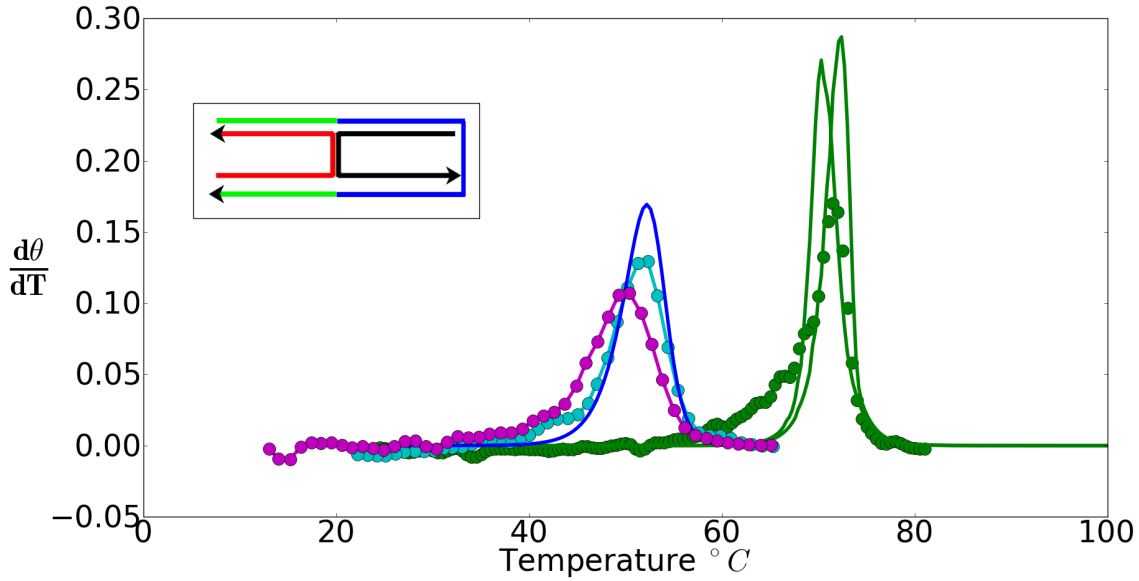


Figure 5: Melting curves obtained for a small origami containing two staples (configuration represented in the inset). Comparison between the experimental (dotted line) derivative $d\theta/dT$ and the modeling (continuous line) as provided by the contribution $\Delta G_{top}(\text{LIS2})$. The curves clustered around $T_m \sim 50^\circ\text{C}$ correspond to B1 in position outer (cyan), B1 in position inner (magenta), the modelling does not make any difference between these two cases. The curves clustered around $T_m \sim 70^\circ\text{C}$ correspond to the simultaneous binding of B1m (inner) and B2m (outer).

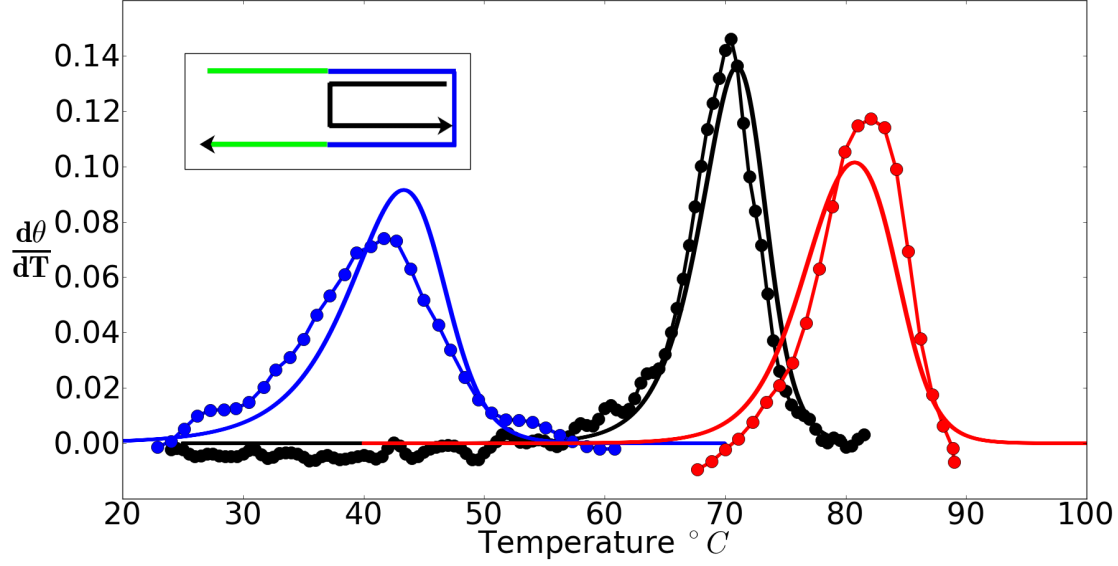


Figure 6: Melting curves obtained for a small origami containing only one staple in the inner position (inset). Comparison between the experimental (dotted line) derivative $d\theta/dT$ and the modeling (continuous line) as provided by the contribution $\Delta G_{top}(LIS3)$. In order of increasing melting temperature: B1 staple (blue), B2m staple (black), B2 staple (red).

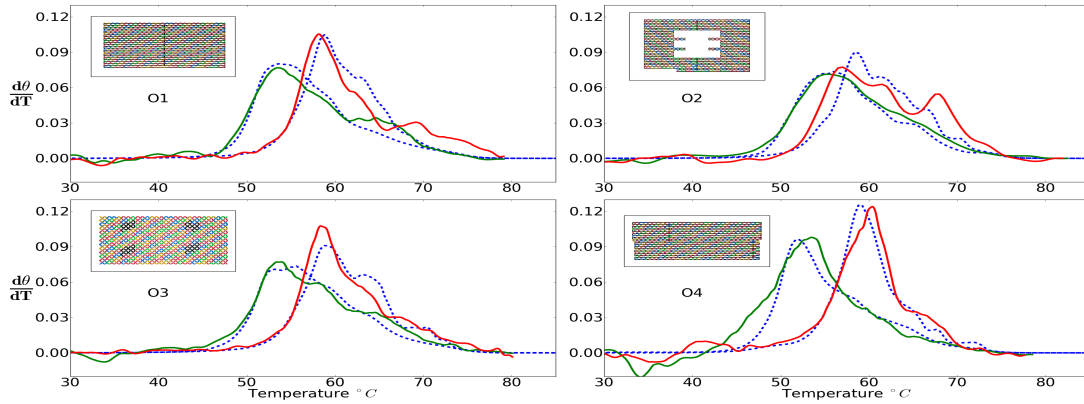


Figure 7: Derivative of the pairing degree versus temperature for the O1, O2, O3 and O4 origamis represented in the insets. The data corresponding to annealing are in red, melting in green, the model (dashed line) is in blue for both processes.

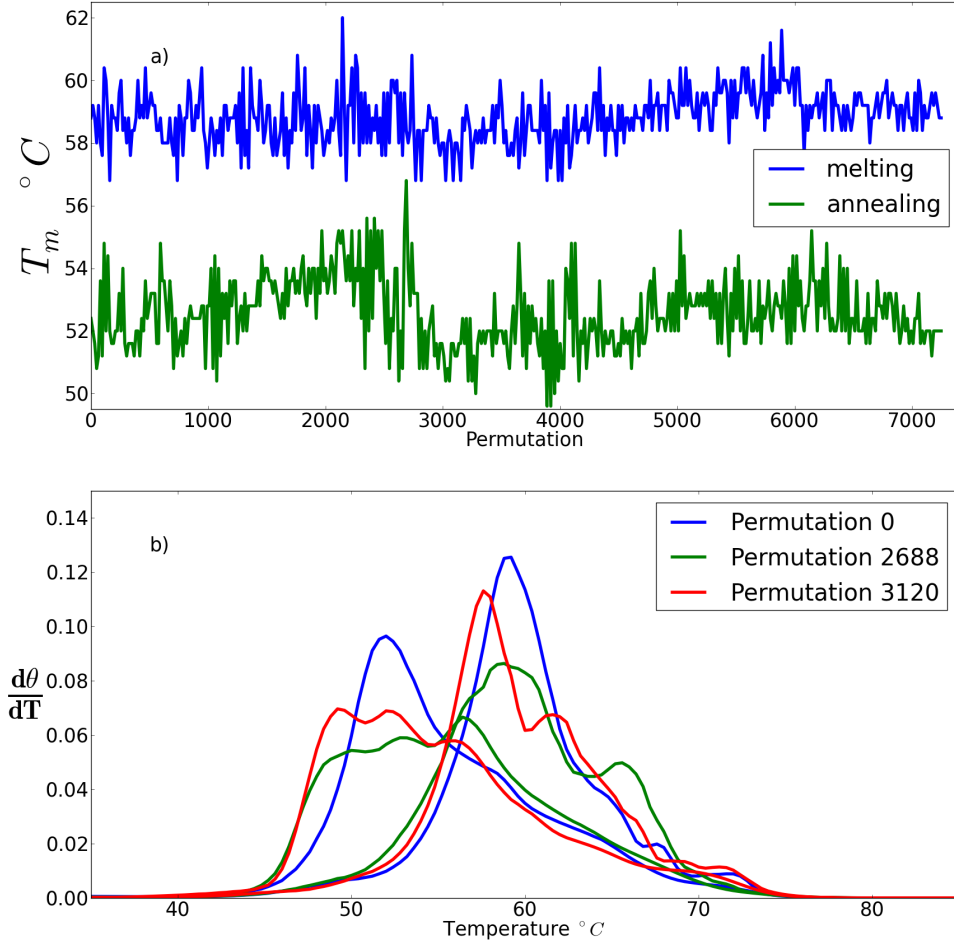


Figure 8: (a) Distribution of melting temperatures (annealing and melting) as a function of the order of the circular permutation of the scaffold strand. (b) two different melting curves corresponding to two permutations with the lowest and highest annealing temperatures.

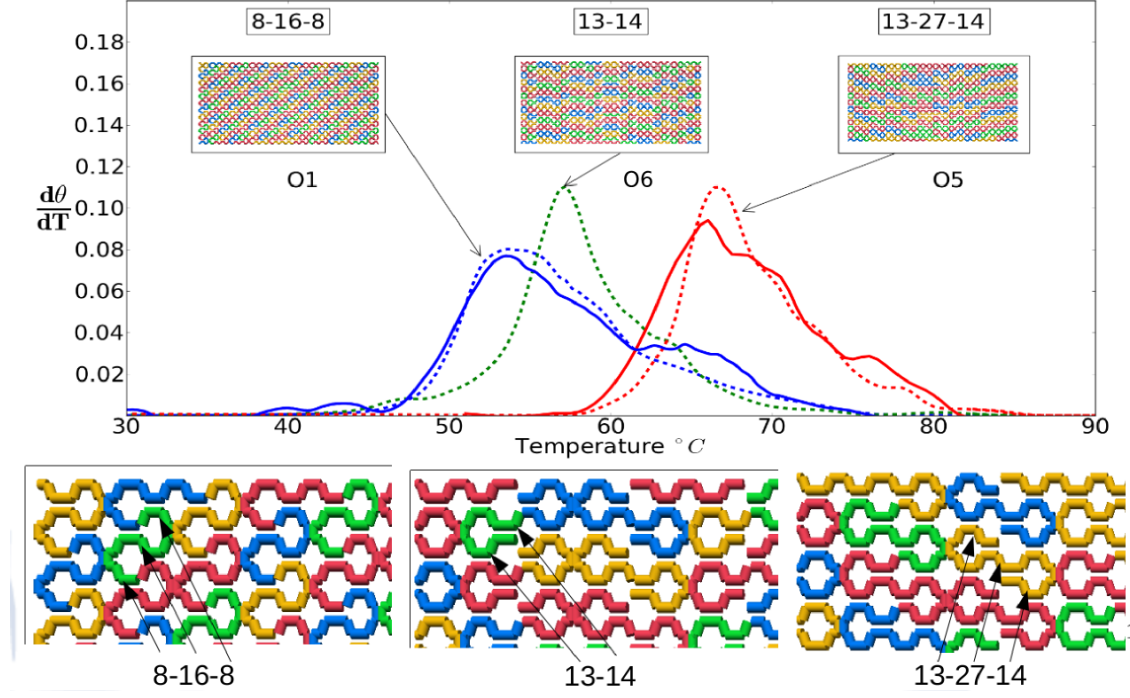


Figure 9: Annealing curves of the O1, O5 and O6 origamis. The three origamis correspond to the same scaffold pattern, but different staple pattern (solid line = experimental data, dashed line = theoretical curves). The lower panel represents a detailed view of the differences in connectivity between the three origamis.

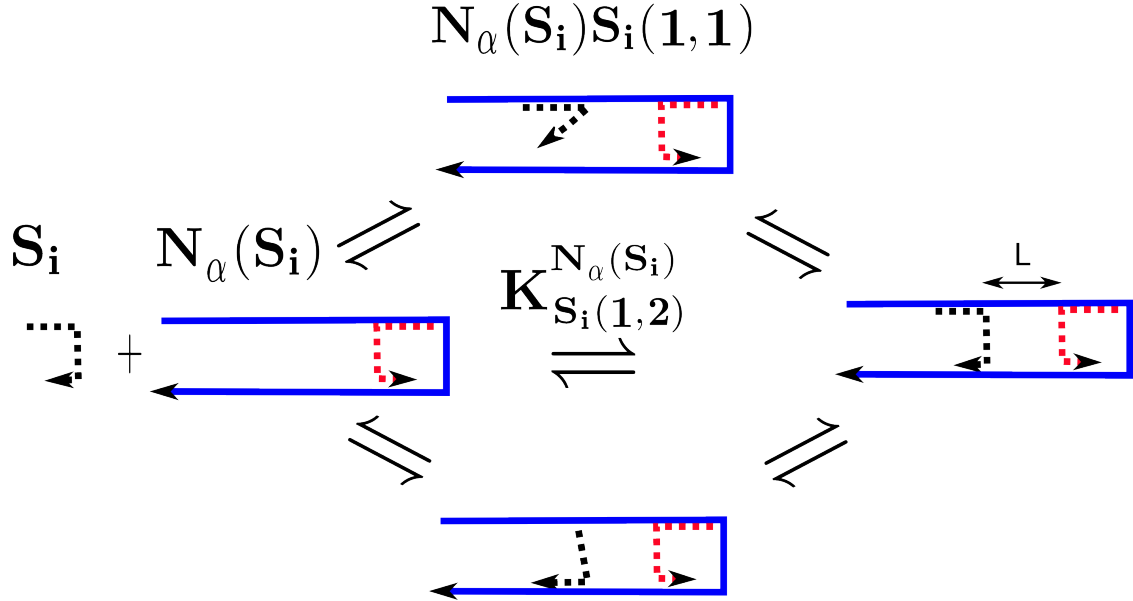


Figure 10: Different possibilities for the staple S_i to bind in the neighbourhood $N_{\alpha}(S_i)$

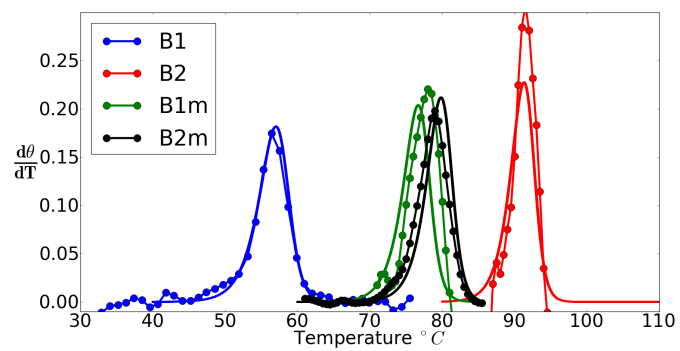


Figure 11: (dotted line = experiment, solid line = theoretical) Comparison between the model and the melting experiments for four different dsDNA